Arunachalam, V.
Evaluation of Diallel Crosses by Graphical and Combining Ability Methods.
EVALUATION OF DIALLEL CROSSES BY GRAPHICAL AND COMBINING ABILITY METHODS

V. Arunachalam

Division of Genetics, Indian Agricultural Research Institute, New Delhi 110 012

(Accepted: 22-vii-76)

The development of a plant breeding strategy hinges mainly on the support provided by genetic information on the inheritance and behaviour of major quantitative characters associated with yield, quality or for that matter, any economic trait of concern to the breeder. To derive such genetic information, it is necessary to conceive of a genetic model in relation to the material that is proposed to be utilised. This process will, in turn, involve in most cases, the designing of a suitable mating system to fit into the chosen genetic model. Diallel crossing is one such important mating system enjoying universal application in plant breeding. However, this mating design can fit into, at least, two genetic models—one essentially developed for a single diallelic gene, but extended, as shall be seen later, to many genes only under strict restrictions (Jinks, 1954; Hayman, 1954, 1957, 1958, 1960; Mather and Jinks, 1971) and the other, a more general model (Griffing, 1956). A lack of appreciation of the basic features and differences of the two genetic models underlying diallel analysis often leads to logical fallacies and errors of interpretation. Moreover, attempts to collate analyses based on both these models can make the confusion worse confounded. It was felt worthwhile therefore, to critically assess these two genetic models especially from the point of view of their utility in plant breeding.

The Models

Model 1: The genetic basis and analysis of this model are fully detailed in Jinks (1954) and Hayman (1954, 1957, 1958, 1960). This model operates under the following major assumptions: (i) parental homozygosity (ii) normal diploid segregation (iii) no difference between reciprocal crosses (iv) no multiple alleles (v) no linkage (vi) no non-allelic gene interaction (vii) genes independently distributed in their effects on parents.

It is apparent that conformity with these severe restrictions will hardly be realised in practice in respect of quantitative characters of a population. In effect, the model reduces to that of a single diallelic gene with diallel mating between the genotypes TT and tt only (where T, t are the alleles at the single locus), mating with and between heterozygotes not being considered (Table 85 in Mather and Jinks, 1971). If the gene frequencies of T and t are p and q
and if the phenotypic values of the genotypes, $TT$, $Tt$ and $tt$ are $d$, $h$ and $-d$ (on an appropriately defined scale, see chapter 4 in Mather and Jinks, 1971), one can obtain, after going through the genetic algebra and ignoring the effects of environment (Hayman, 1954; Mather and Jinks, 1971), the following statistics to provide estimates of the genetic components of variation.

\[ V_p = D; \quad V_r = \frac{1}{4}(D + H_1 - H_2 - F); \quad V_i = \frac{1}{4} H_2; \quad W_r = D - \frac{1}{4} F, \]

the notations being the same as given by Mather and Jinks (1971). Estimates of $D$, $H_1$, $H_2$ besides graphical analysis of the regression of $Wr$ on $Vr$ using $Vr$--$Wr$ limiting parabola will provide a complete genetic interpretation of results of an actual experiment.

It is pertinent to note that there are no built-in parameters in the genetic model envisaged, to provide for epistasis which is considered to be absent. This situation clearly differs from that in a genetic model which provides for the existence of epistatic components and enunciates the conditions under which they are zero. For purposes of reference, we designate the analysis under this model as Graphical Analysis (G.A.).

**Model 2:** The complete specification, analysis and interpretation of this model are given by Griffing (1956). We designate this method of analysis as combining ability analysis (C.A.).

In essence, in this model the metric value of a cross is partitioned into two major components, apart from a general mean and an environmental component, — (i) the contribution of the parents, the general combining ability ($gea$) effect, analogous to a main effect of a factorial design and (ii) the excess over and above the sum of the two $gea$ effects, called the specific combining ability ($sea$) effect, analogous to an interaction effect of a factorial design.

This model operates under the sole assumption that parents of the diallel cross are inbred; a major advantage of this is that, when the hybrids and the parents are raised in plots (usually one row or few rows) in the field, genetic uniformity will be maintained so that the quantitative characters can be measured on a random sample of plants from a plot.

A particular genetic constitution of parents, like the one-gene diploid homozygous genotypes, $TT$, $tt$ of Model 1, is not assumed in this model. So much so, genetic information pertaining specifically to one gene model such as $D$, $H_1$, $H_2$ etc., would not be available; nevertheless, relevant and vital genetic information of direct utility to plant breeding problems would be provided by this model, as will be seen later.

The singular advantage of this model is that the estimates of the variances due to $gea$ and $sea$ provide an apt diagnosis of the predominant role of additive or epistatic gene action. The $gea$ and $sea$ effects will help to locate the parents and crosses that are responsible in bringing about a particular type of gene action.

In fact, the variances of $gea$ ($\text{Var } G$) and $sea$ ($\text{Var } S$) are derived from the covariance of relatives, particularly from the covariances of full- and half-sibs.

\[
\text{Cov (Half-sibs)} = k\sigma_a^2 + k^2\sigma_{AA}^2 + k^3\sigma_{AAA}^2 + \ldots
\]
Our main purpose is to consider the merits of employing either or both the methods—G. A. and C. A.—in the analysis of data from diallel crosses, particularly from the point of view of the plant breeder.

As outlined earlier, the G. A. is based on a single diallelic gene model. The total genetic variance, being the variance of the genotypic values of the 3 genotypes, has only 2 d.f. which are uniquely accounted for by the additive and dominance variances (Arunachalam and Owen, 1971). A partitioning of the total genetic variance into its additive and dominance components was first given by Fisher (1941). He defined ‘average excess’ $a$, of any measurement in respect of any gene substitution, such as that of $T$ for $t$, as the difference in the mean values of the two moieties into which the population of the three genotypes, $TT$, $Tt$ and $tt$ would be divided. If $p^2$, $2pq$ and $q^2$ are the frequencies and $d$, $h$, $-d$ are the genotypic values of the three genotypes (ignoring environmental effects), the average excess, $a = \frac{p^2d}{p^2 + 2pq} + \frac{pqh}{2pq} - \frac{q^2d}{pq + q^2} = d + (q-p)h$.

The ‘average effect’, $z$, in respect of any gene substitution is the linear regression coefficient obtained by fitting an additive model to the genotypic values. An outline of the additive model fitted to the genotypic values of $TT$, $Tt$ and $tt$ is given below:

<table>
<thead>
<tr>
<th>Genotype</th>
<th>Frequency</th>
<th>Genotypic Value</th>
<th>Addition Value</th>
<th>Dominance Deviation</th>
</tr>
</thead>
<tbody>
<tr>
<td>TT</td>
<td>$p^2$</td>
<td>$d$</td>
<td>$g + 2a$</td>
<td>$d-g-2a$</td>
</tr>
<tr>
<td>Tt</td>
<td>$2pq$</td>
<td>$h$</td>
<td>$g + a$</td>
<td>$h-g-a$</td>
</tr>
<tr>
<td>tt</td>
<td>$q^2$</td>
<td>$-d$</td>
<td>$g$</td>
<td>$-d-g$</td>
</tr>
</tbody>
</table>
We note that we have not assumed any environmental effect so that the
phenotypic values \( d, h \) and \(-d\) represent also the genotypic values. On fitting
the additive model, we get an estimate of \( \alpha \), by least squares method.

\[
\alpha = p(d\text{ (difference in the genotypic values of } TT, tt) + q(h\text{ (difference in the genotypic values of } Tt, tt))
\]

\[
= p(d-h) + q(h+d)
\]

\[
= d + (q-p)h
\]

Thus, we see that, under random mating, \( \alpha = a \) and it can be shown to be equal
to the additive effect, \( L \). The dominance effect, \( Q \) can be shown to be equal to
\(-2h\), so that the dominance deviations of the genotypes \( TT, Tt \) and \( tt \) will
be \( q^2Q, -pqQ, p^2Q \).

Thus, additive effect, \( L = a \quad a = d + (q-p)h \)
dominance effect, \( Q = -2h \)

additive variance, \( VA = 2pqL^2 \) and
dominance variance, \( VD = p^2q^2Q^2 = 4p^2q^2h^2 \) (See for details, Fisher,
1941; Li, 1955; Falconer, 1964; Arunachalam and Owen, 1971).

It is quite clear, therefore, that \( L \), the additive effect is not equal to \( d \), in
general; and \( L = d \) only if \( p = q \). It is essential to note that both \( L \) and \( VA \)
deepend on gene frequencies while \( Q \) does not. In G.A., the component,
\( VA = \frac{1}{2}D_R \) in the notation of Mather and Jinks, 1971) is further split up and
expressed in terms of \( D, H_1, H_2 \) and \( F \) (see chapter 9 in Mather and Jinks, 1971).

In fact, \( VA = 2pqL^2 \)

\[
= 2pq\left[d + (q-p)h\right]^2
\]

\[
= 2pq\left[d^2 + (q-p)^2h^2 + 2dh(q-p)\right]
\]

\[
= 2pqd^2 + 2pq(1-4pq)h^2 + 4pqdh(q-p)
\]

\[
= \frac{1}{2}D_R = \frac{1}{8}(D + H_1 - H_2 - F) \quad \text{(1)}
\]

The terms \( d^2, h^2 \) and \( dh \) can be replaced by \( \Sigma d^2, \Sigma h^2 \) and \( \Sigma dh \) (the respective
values being summed over all the loci), if one deals with a number of purely
additive and completely independent genes (though it is an untenable
assumption), as done by Mather and Jinks (1971).

We notice that \( D \) is not equal to the additive genetic variance, unless
\( p = q \) and that \( VD \), the dominance variance = \( p^2q^2Q^2 = 4p^2q^2h^2 \) has nothing
in common with \( H_1 \) or \( H_2 \) except for the constant, \( h = -\frac{1}{2}Q \). Further, we see
from (1) that \( H_1 = 4pqh^2 \) and \( H_2 = 16p^2q^2h^4 \), so that \( H_1/4H_2 = 1/16pq \)
which will be \( = \frac{1}{4} \) when \( p = q \). Thus \( H_1/4H_2 \) becomes a measure of the
deformation from equality of alleles with positive and negative effects. “This, of
course, is subject to an error depending on the standard errors of \( H_1 \) and \( H_2 \),
which are unknown. Since the estimate of \( pq \) is obtained solely from terms in
\( \Sigma h^2 \), it will, of necessity, only cover the frequencies of allelomorphs of genes exhib-
it degree of dominance. Thus, it provides no evidence about the
distribution of allelomorphic pairs exhibiting no dominance.” (Jinks, 1954).

Moreover, there appears to be no physical or biological value that can be
attached to \( H_1 \) or \( H_2 \) as such.
At this stage, it is important to emphasise that \( L = d + (q-p)h \), is the additive effect always (which, incidentally, \( = d \) if \( p = q \)) and it derives its meaning solely from the fitting of an additive model to the genotypic values, as mentioned earlier. Recollecting that the estimated value of \( Q \), as a consequence of fitting the additive model, \( = -2h \) (while \( h \) is the value of the heterozygote actually measurable), we see that \( L = d + \frac{1}{2}Q(p-q) \). Thus, the additive effect and hence the additive variance \( VA \) or the value of \( DR ( = \Sigma 4pqL^2) \) contains some effect of \( h \), and the dominance variance \( VD \) or the value of \( HR ( = \Sigma 16p^2q^2 h^2) \) is correspondingly less than the summed effects of all the squared \( h \) deviations, under random mating, if \( p \) is not equal to \( q \). "In general, therefore, \( DR \) is not the additive variation.... It is, in fact, the additive variation in the statistical sense, rather than in the genetical sense (Mather and Jinks have adopted, see pp. 213–214 in Mather and Jinks, 1971)." The logic behind the last statement is too weak to justify it. We should remember that, without a statistical fitting of an additive model, an additive effect cannot be defined and obtained. Further, there are no two definitions—one statistical and the other genetical—of an additive effect. The only definition of an additive effect is given by Fisher (1941) as mentioned earlier. Hence the categorisation of a statistical and genetical additive variance stands unwarranted. Thus we see, as explained earlier in discussion, \( \Sigma pqL^2 \) is the additive variance, where \( L = d + (q-p)h \) and since \( d \) is not equal to additive effect, in general, it is illogical to argue that \( DR \) is not the additive variance in the genetical sense. In fact, \( DR \) is the correct expression for additive variance under random mating and it contains terms relating only to additive effects.

Turning now our attention to plant breeding, it is now known that almost all quantitative characters that are dealt with there are governed by more than one gene. In the development of G.A., a one-gene model is assumed to start with and the analysis then extended to many genes under almost impossible restrictions. The total genetic variance, having only 2 d.f., is uniquely partitioned into additive and dominance variances in a one-gene model, as explained earlier. Hence, the extended method using this model can, at the most, estimate (not necessarily unbiasedly) only these two variances. This is obviously inadequate, when it is known that the number of d.f. for the total genetic variance is more than two and that epistatic interactions play an important role.

On the other hand, removing epistatic interactions in actual experimental data by scaling does not appear to be a viable proposition, since it would not be possible to take the interpretation of all the results back to the original data, any longer. According to Mather and Jinks (1971, p. 63), "We cannot even assume without evidence that a scale appropriate to the representation of variation of a character in one set of individuals under one set of conditions will be equally appropriate to the representation of that same character either in a different set of individuals, which may be heterogenic for different genes, or under different conditions. It may well, therefore, never be possible to construct an a priori scale for the representation of variation in a character. Certainly with no more
than our present knowledge of gene action the construction of such a scale is impossible."

For a plant breeder, necessarily working with a number of filial generations, these observations would, therefore, imply that the results on scaled data would not, in a majority of cases, be of any direct value for plant breeding. Similarly, extension of the concepts of G.A. to two genes with linkage and other cases (Mather and Jinks, 1971) is constrained by the restrictions carried forward from one gene theory and offers very limited scope for direct utility in plant breeding programmes.

Another capability claimed for G.A. is the estimation of the effective number of factors controlling a character. The process of estimation involves a number of assumptions and hence the estimate of the effective number of factors is subject to severe limitations from utility point of view (as explained by Mather and Jinks, 1971; Falconer, 1964). Further, the breeding methods that are available do not rely on the number of effective factors controlling a character. To be a result of basic importance, the method of estimation of effective factors should be so sound as to get stable, unbiased and repeatable estimates. It is thus seen that little practical significance can, if at all, be attached to the estimated number of effective factors. By similar arguments, it would easily be seen that other information provided by G.A. like the distribution of alleles with positive and negative effects, the dominance or otherwise of increasing or decreasing alleles in a majority of loci, will be of little utility, if at all, in practical plant breeding.

On the other hand, C.A. scores over G.A. in many respects: (i) It is not restricted to one gene, operates with limited feasible assumptions and hence is more realistic to plant breeding problems. (ii) The gca and sca effects and variances are very effective genetic parameters of direct utility to decide the next phase of the breeding programme. (iii) It enables a plant breeder to decide, for example, about hybrid or pure line breeding, the choice of parents for construction of synthetics, the selection of suitable F$_1$'s for a multiple crossing or a composite breeding programme and the possibility of employing appropriate selection techniques like modified mass selection, recurrent and reciprocal recurrent selection etc. (iv) The designs of experiments associated with C.A. (especially the fixed effects model) are of direct relevance to plant breeding and permit efficient estimation of environmental variance at plot and plant level. (v) The combining ability estimates provide valuable clues to the choice of parents for a multiple crossing programme and to the expected level of heterosis (details under publication) (vi) By employing suitable designs like triallel and appropriate systems of four way crosses, it is possible to get valid estimates of the components of epistatic interaction. (vii) The direction of combining ability effects provides useful inference of value to plant breeding. For example, the probability of obtaining heterosis in a single cross, when the divergence in the direction of gca effects of parents is ensured, is found to be high when the sca effect is not significant (details under publication).
These ideas will gain support, if now we examine some published results where data on diallel crosses have been examined both by G.A. and C.A. There is a set of papers—set A—proclaiming that both the methods yield comparable and almost the same results (see, for example, Nagur and Murthy, 1970; Rao, 1970; Jones, Gupton and Terrill, 1972). There is a host of others—set B—in which the two methods give different results (see for example, Gill, Dhillon and Bains, 1972; Tandon, Joshi and Jain, 1970; Paroda and Joshi, 1970). Most of the cited papers deal with self-pollinated crops where the probability of the assumptions required by G.A. to be true is high. However, a theoretical analogy of their results points out that, while set A cannot prove the validity of G.A., set B does prove the questionable reality of assumptions and doubtful utility of the results in plant breeding.

A view is prevalent among breeders that G.A. provides results on gene action and C.A. on combining ability; some observe that the former is a genetical and the latter a statistical approach. In the context of what we have discussed in this paper, it is obvious that this categorisation is unwarranted. For example, as seen earlier, the term additive genetic effect derives its very meaning from a statistical approach. Hence this arbitrary categorisation should not be the reason why both the methods should be applied to the same data. As a matter of fact, one set (or a few sets) of data from field experiments (however well they might have been conducted) cannot provide the base to evaluate the validity of a one gene as well as of a more general model for practical breeding problems. Further, our aim in plant breeding is not to test whether a one gene model fits a particular set of data (though, in some special circumstances, it may be warranted) but to use a model which rests on realistic and bread-based assumptions and which provides useful clues to breeding. Hence, the realistic choice of a model should be done at the beginning of an experiment. The technique of subjecting the data to a number of methods that are and will be available (ignoring the assumptions and the genetic models inherent in them) and collating the various results obtained as done by breeders, is definitely not a scientifically sound procedure (see for example, Paroda and Joshi, 1970). A loud scientific thinking suggests that C.A. is more relevant to plant breeding experiments than G.A.

During the examination of a number of published papers, the following deficiencies in the design of experiments or interpretation of diallel crosses have been observed.

(i) Extension of C.A. to F2 data needs considerable precautions especially in view of the fact that the segregation observed in different crosses is, as a rule, not comparable. Further, a large population of each F2 should be grown and a large enough sample per plot should be used to get, if at all, any meaningful results. C.A. is a testing procedure to choose parents/crosses in the earliest generation to obtain the next generation. Hence advocating the C.A. in F2 and higher generations, on the basis that adequate seeds in F1 is a limiting factor, is questionable in view of a differential segregation in F2 (which confers some
unknown advantage to certain cross combinations) and, when a complete genetic homogeneity is not certain, an unconscious selection in F₁ (due to the differing number of seeds in various crosses). When seeds are limited in F₁, it is useful to repeat the crosses in an off-season, pool up sufficient seeds and then test F₁. The associated time lag is inevitable if the F₁ testing is to be of any significance.

(ii) Sometimes, combinations of some generations of diallel cross material (like F₁, F₂, F₃, backcrosses etc) are raised in one randomised block design with different weights on the generations (like plots of 1 row for F₁, 3 rows for F₂, etc). The C.A. is done on these generations separately pulling out only the data on relevant entries from the r.b.d. (see, for example, Tandon, Joshi and Jain, 1970; Singh and Singh, 1972; Singh and Dhaliwal, 1972; Dhaliwal and Gill, 1973; Singh and Singh, 1974). This leads to different estimates of error and other fallacies (see Arunachalam, 1974) and should be avoided. It is easier to plant the material of different generations in small r.b.d.’s in contiguous blocks of experimental area.

(iii) A number of papers is found in which the fixed effects model is used and the estimates of gca, sca variances are calculated on random effects model (see, for example, Gill, Dhillon and Bains, 1972). This is obviously erroneous.

(iv) The magnitude of additive as compared to non-additive variance can be inferred only by a comparison of the estimates of gca and sca variances obtained from the expected values of gca and sca mean squares (m.s.) in the combining ability ANOVA. Unfortunately, in several published papers, the magnitudes of gca and sca means have been compared instead, to decide the predominance of additive to non-additive gene action (see, for example, Kronstad and Foote, 1974; Paroda and Joshi, 1970; Singh and Singh, 1974; Dhaliwal and Singh, 1973; Murty, 1975).

(v) An attempt to draw relevant genetic information on yield should always be made on a comparative examination of all the yield components taken together, in preference to an examination of a subset of components in isolation (unlike the papers by Singh and Singh, 1972 on 3 characters and by the same authors in 1974 on 5 characters on the same experiment in mung bean).

(vi) It will always be helpful if a breeder, where possible, ensures the reliability of the results on gene action obtained from F₁ by a follow-up of selected material in later generations. For example, a highly heterotic cross can be repeated and crosschecked for the degree of heterosis; a significant additive genetic variance found in a particular population can be inferred by measuring the response to selection in the immediately succeeding generation. Results reported on such cross-checks will not only be of scientific merit but also render repeatability, an essential component of breeding methods, feasible.

Thus, it would appear from the analysis in the foregoing pages, that all aspects of relevant genetic information can profitably be obtained by C.A. alone, in preference to the analysis by G.A., for plant breeding problems.
SUMMARY

The genetic models behind two approaches, graphical analysis due to Jinks and Hayman and combining ability analysis due to Griffing, employed in the analysis of data from diallel crosses are reviewed and their relative merits in plant breeding discussed. It is shown that the methods given by Griffing provide all the information that a breeder will need from a diallel cross. A few lapses that were noticed in published literature in the diallel cross experiments are brought out.

ACKNOWLEDGEMENT

The author is grateful to Professor S. Ramanujam and Dr. A. Bandyopadhyay for illuminating discussions which helped to improve the manuscript.

REFERENCES


